SRI International

CCNY-SRI: An interactive visual event detection system

Chenyang Zhang*, Xiaodong Yang*, Chucai Yi*, Yingli Tian*,
Qian Yu**, Amir Tamrakar**, and Ajay Divakaran**
*The City College of New York,
** SRI International Sarnoff

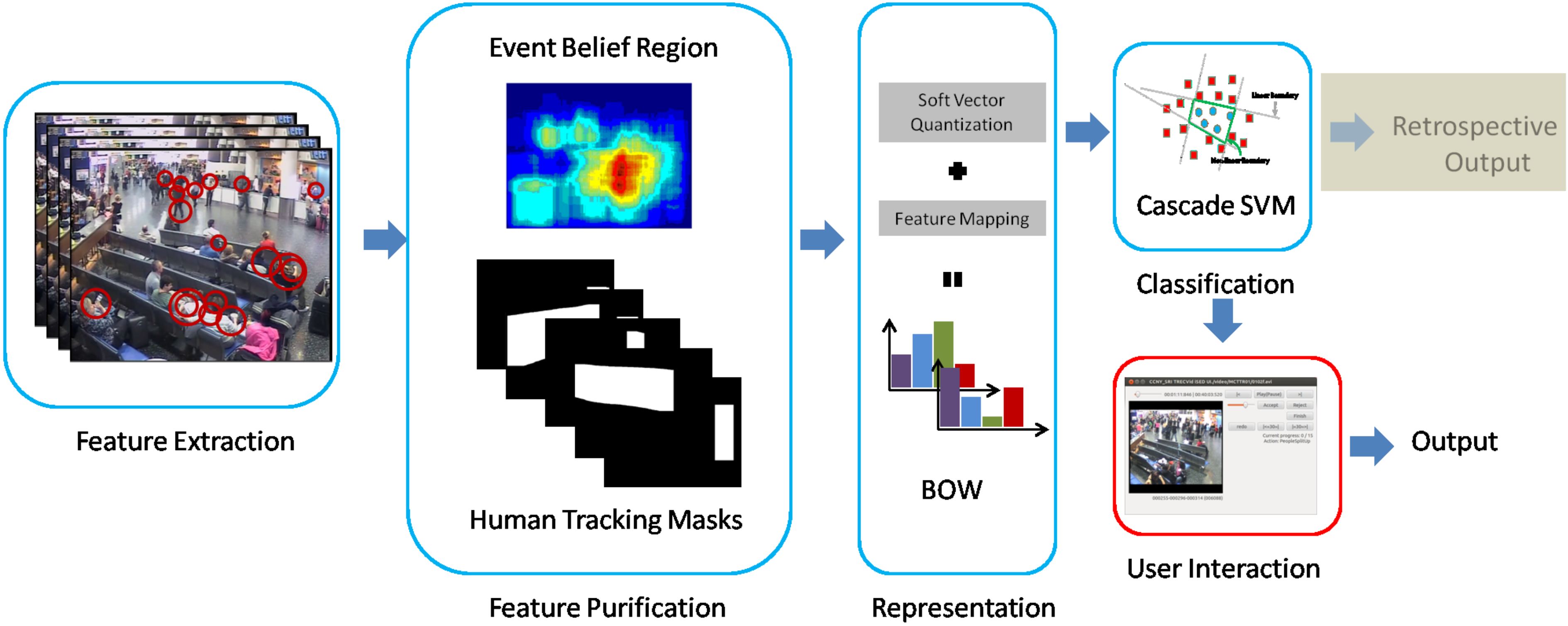
About Us

- Media lab, The City College of New York (CCNY)
- SRI International

SRI International

- We participated last year’s SED task as ``MediaCCNY’’ for the 1st year

Overview of Our System



Feature Extraction

Event Belief Region

Human Tracking Masks

Feature Purification

Soft Vector Quantization

+

Feature Mapping

=

BOW

Representation

Cascade SVM

Classification

Retrospective Output

User Interaction

Output

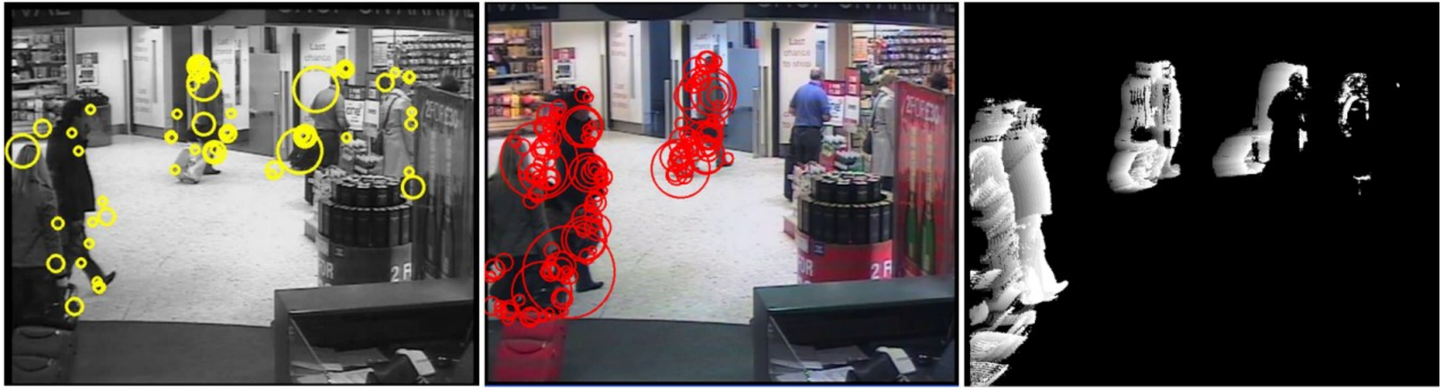
- Human tracking is involved
- User is involved as the final decision maker

Outline

- Feature Extraction
- Feature Purification
- Representation
- Event Inference (Classification)
- User Interaction

Feature Extraction

- 2 feature channels are used:
  - 1. STIP-HOG/HOF
  - 2. SURF/MHI – HOG



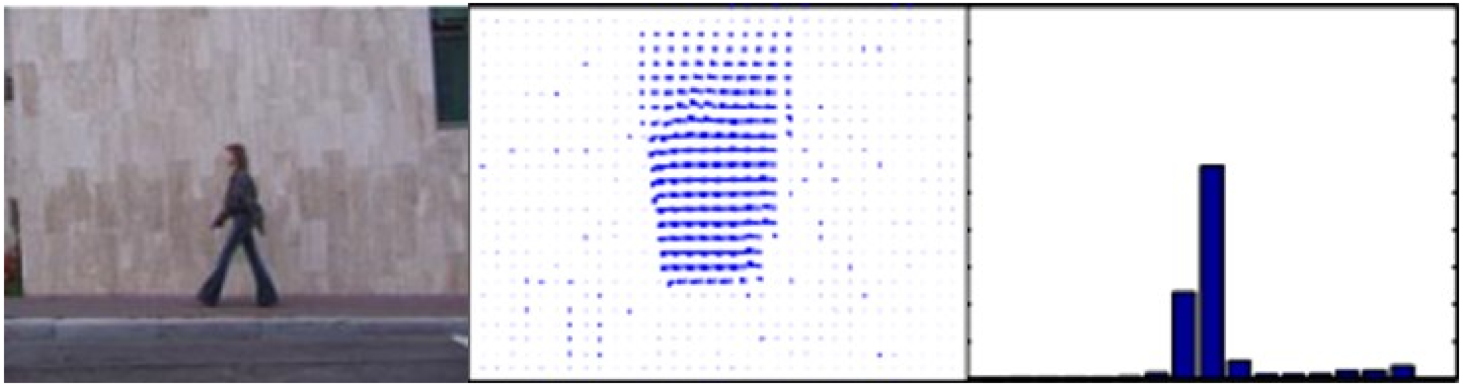
STIP SURF/MHI Motion History Image

  - Two detectors extract complementary interest feature points
  - Frames are downsampled: 720x576 -> 360x288

Feature Extraction

- Descriptor Channels:
  - Histogram of Gradients (HOG) Spatial Feature

  - Histogram of Flows (HOF) Temporal



Outline

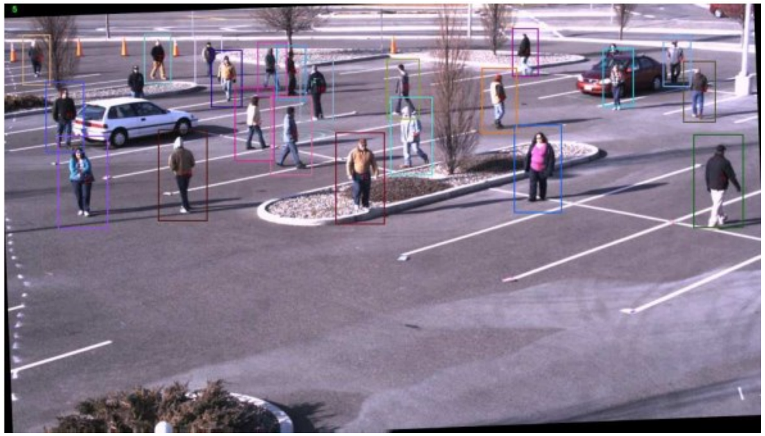
- Feature Extraction
- Feature Purification
- Representation
- Event Inference (Classification)
- User Interaction

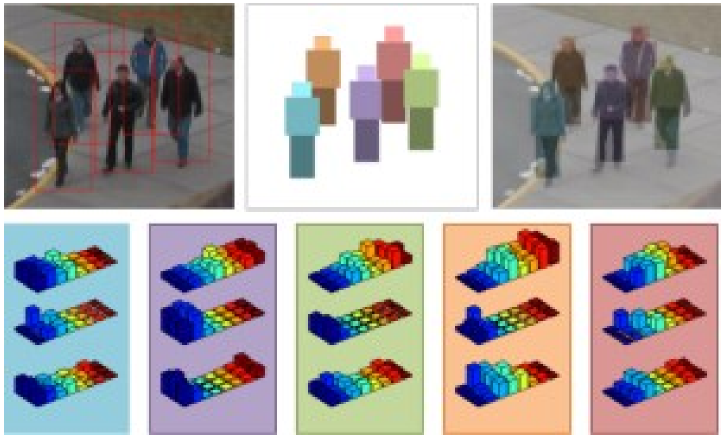
Feature Purification

- Two issues with extracted feature points:
  - Huge number
  - Too much Noise

- Feature purification is conducted on:
  - Objective Saliency Capture (moving people)
  - Semantic Saliency Capture (event frequency prior)

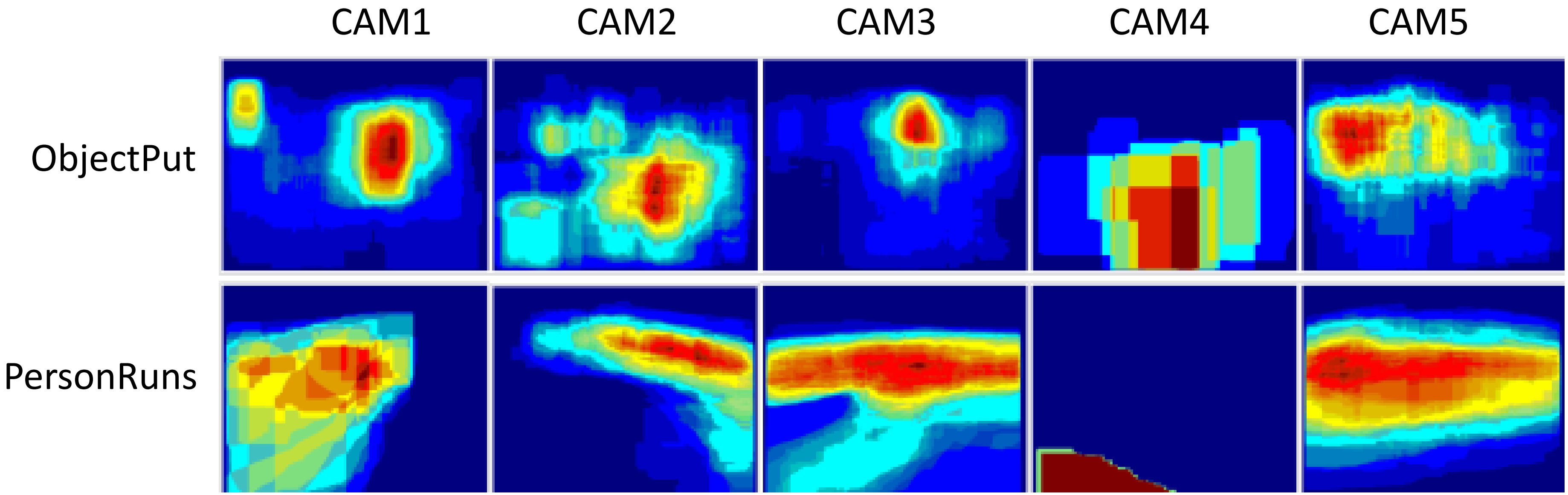
Human Tracking Mask





- Multiple human tracking bounding boxes are used as filtering masks

Event Belief Region

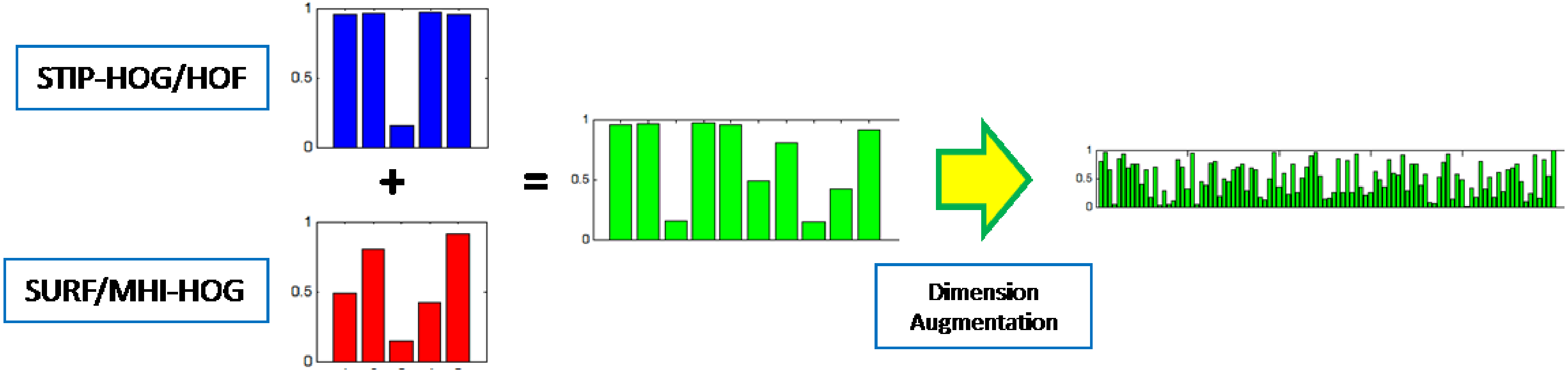


- Event specific event belief region is used to capture semantic saliency

Outline

- Feature Extraction
- Feature Purification
- Representation
- Event Inference (Classification)
- User Interaction

Feature Representation



STIP-HOG/HOF

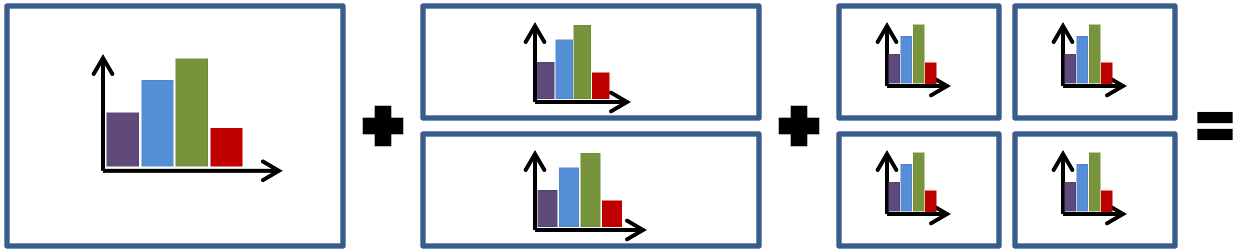
SURF/MHI-HOG

Dimension Augmentation

- Local features (short strings) inside a ``window” are aggregated using Bag-of-words model
- Dimension Augmentation using feature mapping (long strings)

Feature Aggregation

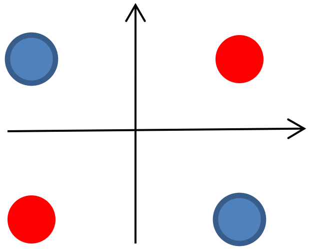
- Feature dimension:
  - STIP-HOG/HOF: 162 SURF/MHI-HOG: 256
- Code book is built on K-means clustering
- Spatial pooling uses a 3-layer pyramid:



Feature dimension: 7K

Feature Mapping

- “XOR” problem:



label	Original feature (x,y)	Mapped feature (x, y, xy)
-1	(1,1)	(1,1,1)
-1	(-1, -1)	(-1,-1,1)
1	(1, -1)	(1, -1, -1)
1	(-1, 1)	(1, -1, -1)

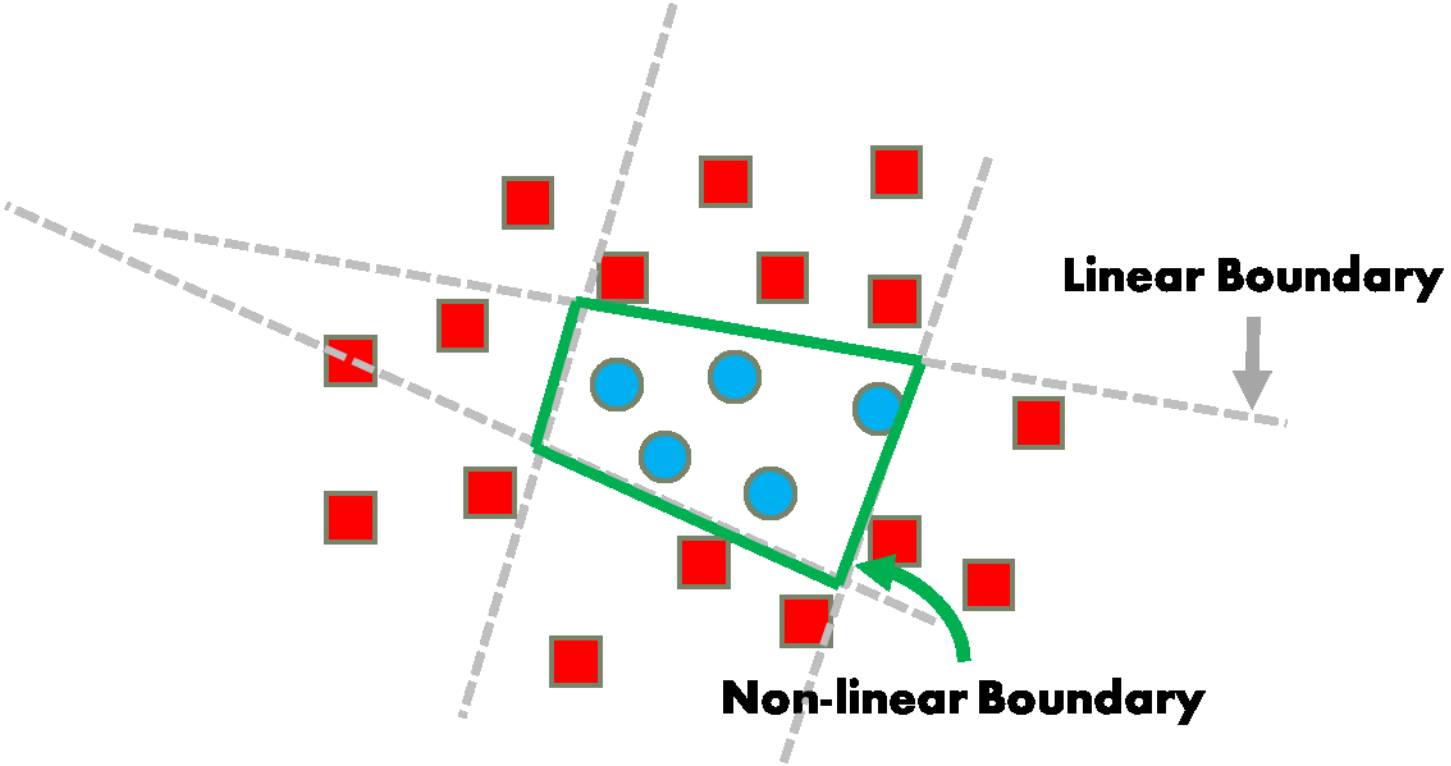
- Feature mapping: map original feature to some high dimensional feature space

Outline

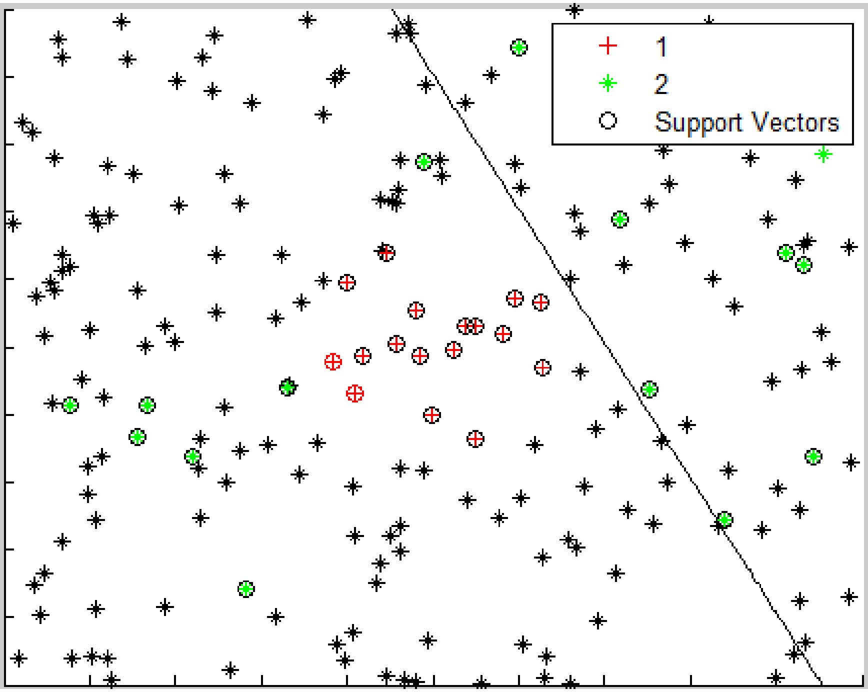
- Feature Extraction
- Feature Purification
- Representation
- Event Inference (Classification)
- User Interaction

Event Inference

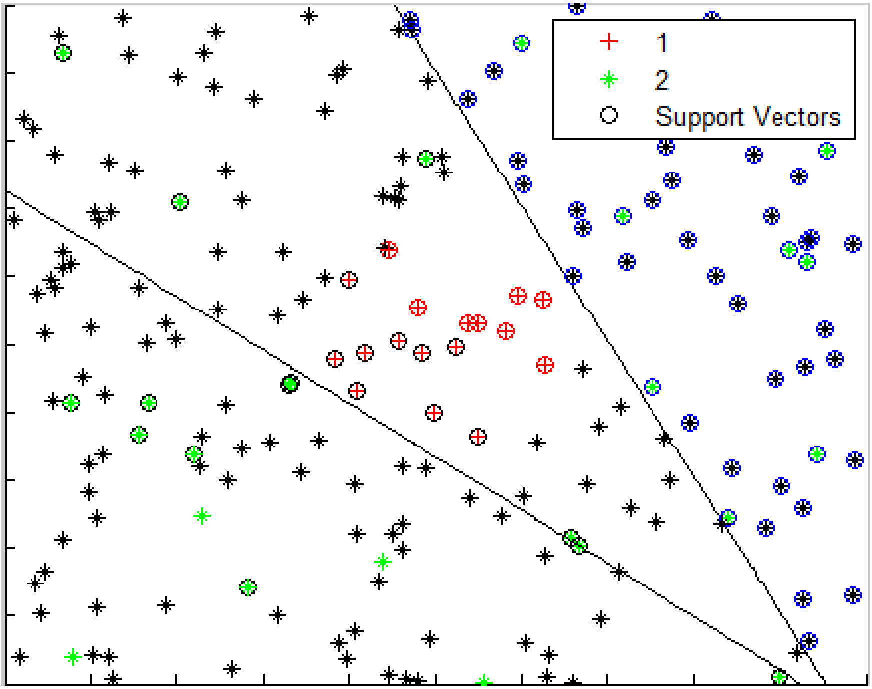
- Cascade SVMs are used as classifier
- Each unit sample is a temporal window of 60



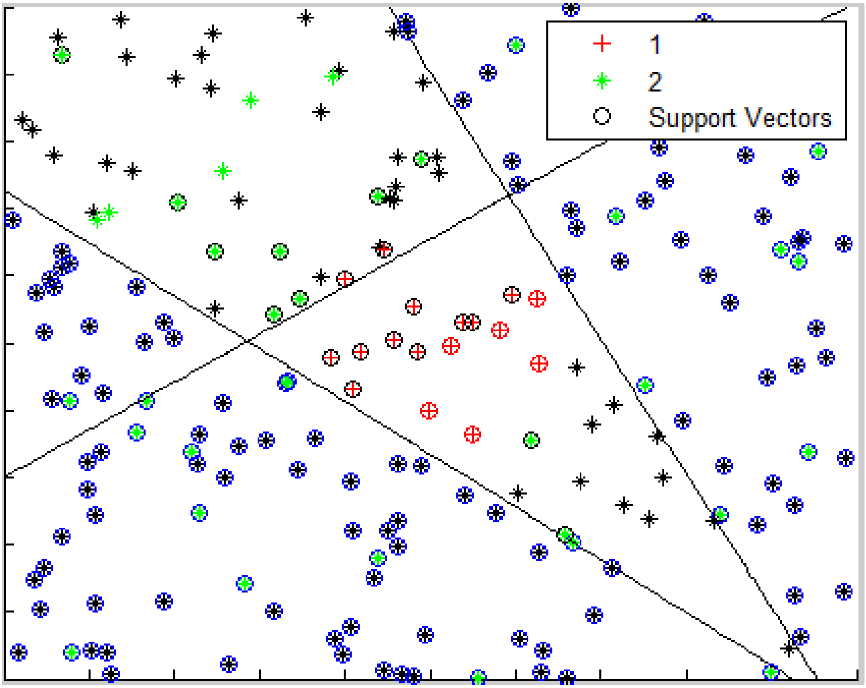
A Demo iter 1



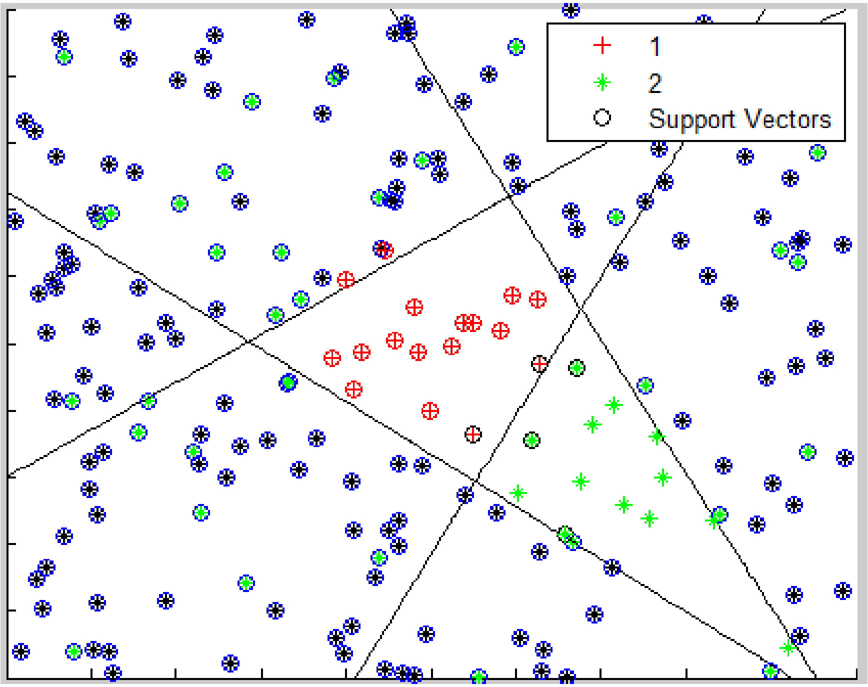
A Demo iter 2



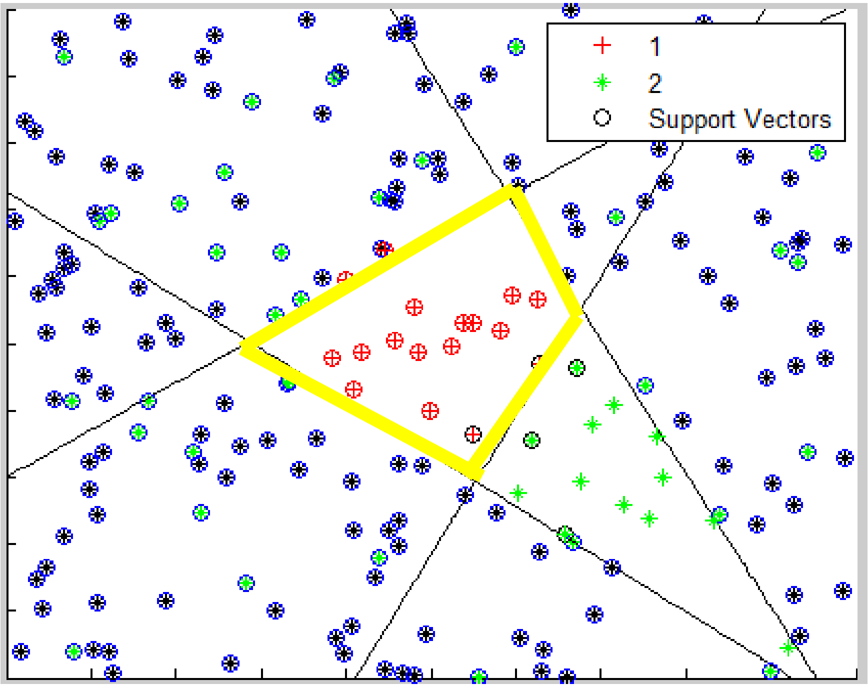
A Demo iter 3



A Demo iter 4



A Demo iter 4

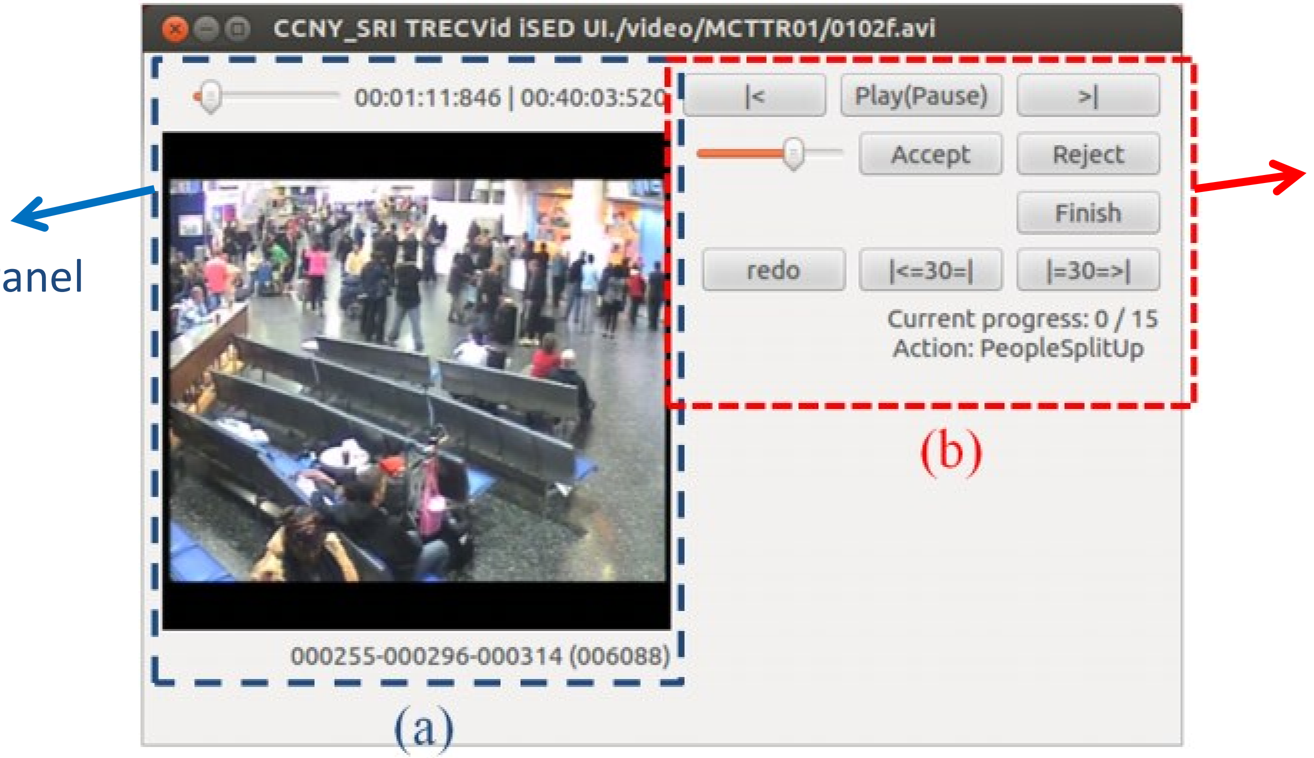


Outline

- Feature Extraction
- Feature Purification
- Representation
- Event Inference (Classification)
- User Interaction

Human Interaction

- Motivation
  - Let an expert user be the final decision maker



CCNY_SRI TRECVid iSED UI./video/MCTTR01/0102f.avi

00:01:11:846 | 00:40:03:520

Current progress: 0 / 15
Action: PeopleSplitUp

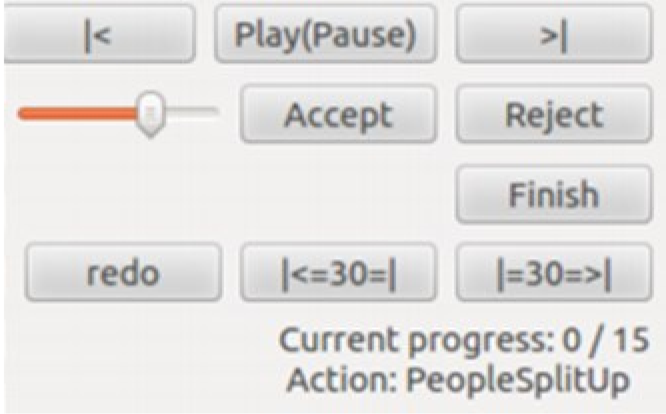
000255-000296-000314 (006088)

Video Panel

Control Panel

Human Interaction

- Some Facts about our UI

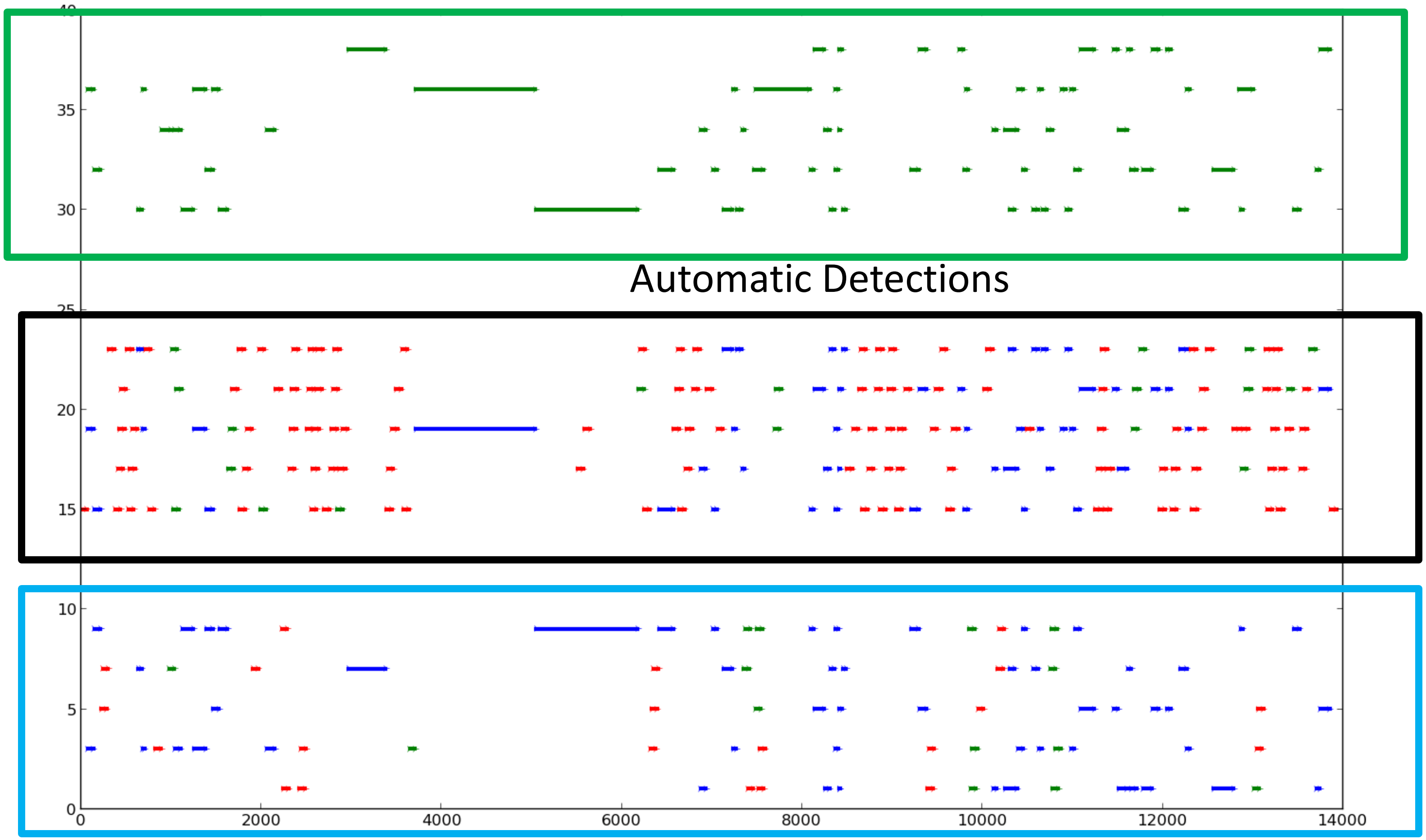


Current progress: 0 / 15
Action: PeopleSplitUp

  - “Reject” is the basic move
  - “<=“ or “=>” are seldom used
  - More than 5 basic moves can be distracting

What did a user do?

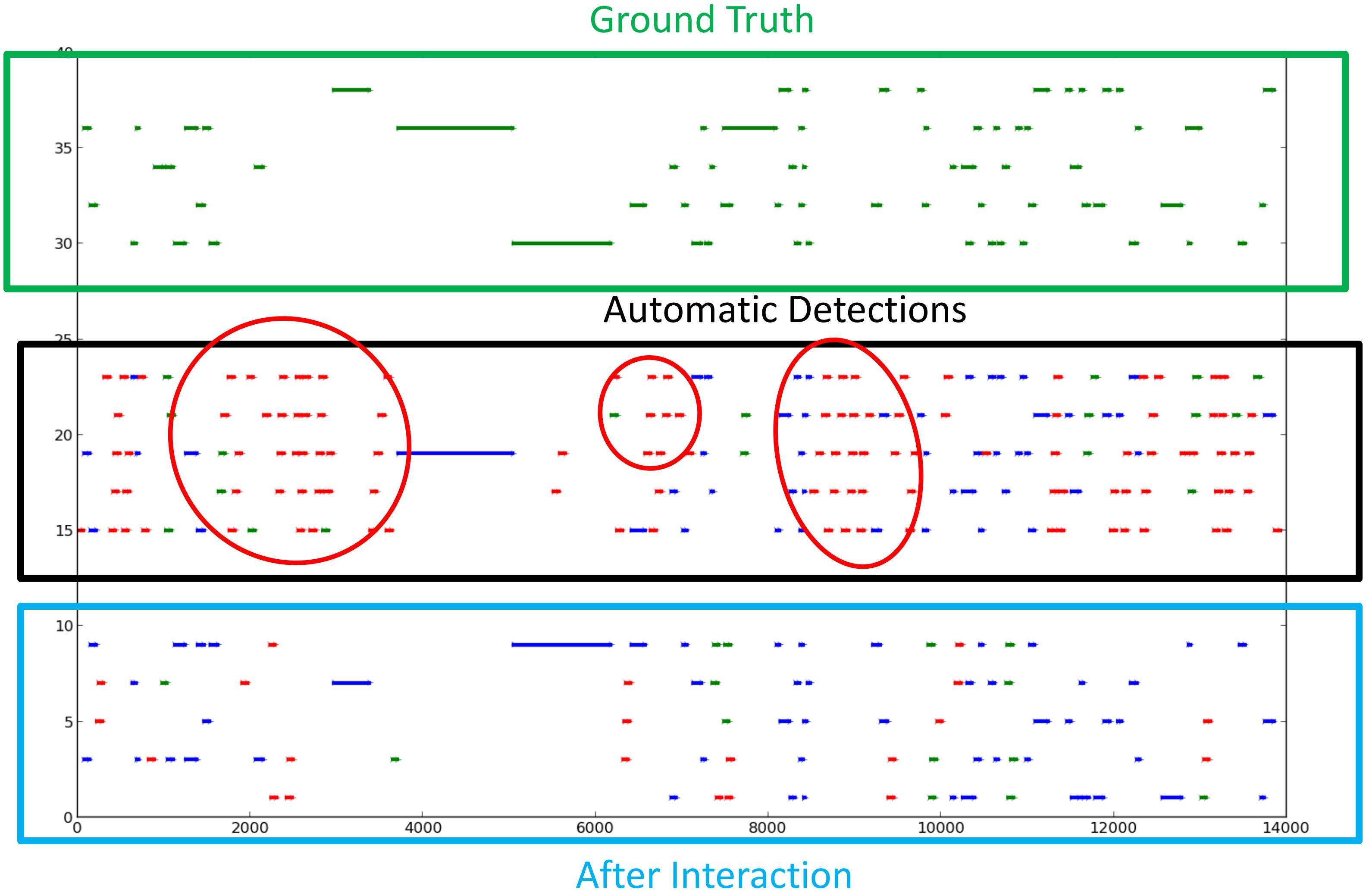
Ground Truth



Automatic Detections

After Interaction

What did a user do?



Results

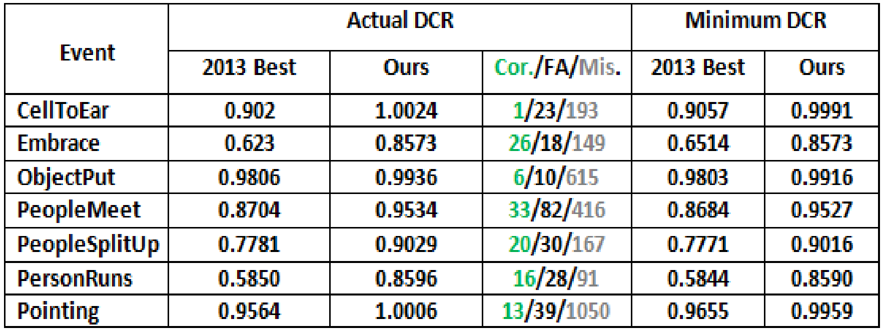
– With 25 mins limit: (rejecting all others)

Event	Actual DCR			Minimum DCR	
	2013 Best	Ours	Cor./FA/Mis.	2013 Best	Ours
CellToEar	0.902	1.0024	1/23/193	0.9057	0.9991
Embrace	0.623	0.8573	26/18/149	0.6514	0.8573
ObjectPut	0.9806	0.9936	6/10/615	0.9803	0.9916
PeopleMeet	0.8704	0.9534	33/82/416	0.8684	0.9527
PeopleSplitUp	0.7781	0.9029	20/30/167	0.7771	0.9016
PersonRuns	0.5850	0.8596	16/28/91	0.5844	0.8590
Pointing	0.9564	1.0006	13/39/1050	0.9655	0.9959

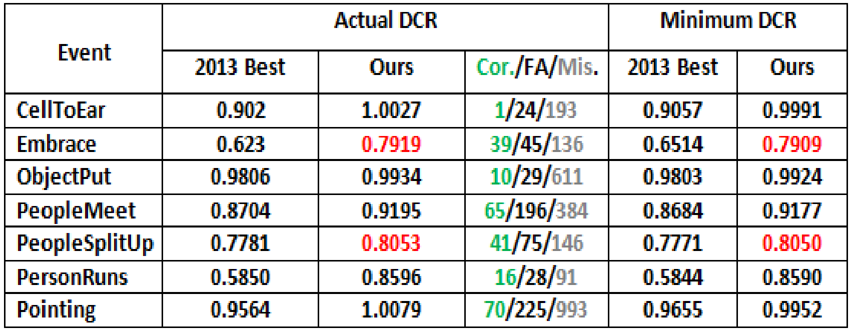
– Remove 25 mins limit:

Event	Actual DCR			Minimum DCR	
	2013 Best	Ours	Cor./FA/Mis.	2013 Best	Ours
CellToEar	0.902	1.0027	1/24/193	0.9057	0.9991
Embrace	0.623	0.7919	39/45/136	0.6514	0.7909
ObjectPut	0.9806	0.9934	10/29/611	0.9803	0.9924
PeopleMeet	0.8704	0.9195	65/196/384	0.8684	0.9177
PeopleSplitUp	0.7781	0.8053	41/75/146	0.7771	0.8050
PersonRuns	0.5850	0.8596	16/28/91	0.5844	0.8590
Pointing	0.9564	1.0079	70/225/993	0.9655	0.9952

Observations



Event	Actual DCR			Minimum DCR	
	2013 Best	Ours	Cor./FA/Mis.	2013 Best	Ours
CellToEar	0.902	1.0024	1/23/193	0.9057	0.9991
Embrace	0.623	0.8573	26/18/149	0.6514	0.8573
ObjectPut	0.9806	0.9936	6/10/615	0.9803	0.9916
PeopleMeet	0.8704	0.9534	33/82/416	0.8684	0.9527
PeopleSplitUp	0.7781	0.9029	20/30/167	0.7771	0.9016
PersonRuns	0.5850	0.8596	16/28/91	0.5844	0.8590
Pointing	0.9564	1.0006	13/39/1050	0.9655	0.9959



Event	Actual DCR			Minimum DCR	
	2013 Best	Ours	Cor./FA/Mis.	2013 Best	Ours
CellToEar	0.902	1.0027	1/24/193	0.9057	0.9991
Embrace	0.623	0.7919	39/45/136	0.6514	0.7909
ObjectPut	0.9806	0.9934	10/29/611	0.9803	0.9924
PeopleMeet	0.8704	0.9195	65/196/384	0.8684	0.9177
PeopleSplitUp	0.7781	0.8053	41/75/146	0.7771	0.8050
PersonRuns	0.5850	0.8596	16/28/91	0.5844	0.8590
Pointing	0.9564	1.0079	70/225/993	0.9655	0.9952

- Significant bias is observed between user judgment and ground truth
  - E.g. in PeopleMeet, user brought in 146 clips, while 114 of them is false alarm.

- Improvement is observed in those events with reasonable number of detections
  - weighted fraction of total time for different events?

Acknowledgement

- Our team members:

Xiaodong Yang

Chucai Yi

Prof. Yingli Tian

CCNY

Dr. Qian Yu

Dr. Amir Tamrakar

Dr. Ajay Divakaran

SRI International

?
Q
&
A
!

Thanks